

# Splotowe sieci neuronowe

Agnieszka Ławrynowicz

2020

# Wprowadzenie

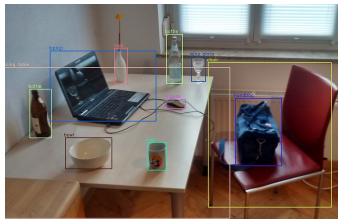
# Zagadnienia rozpoznawania obrazów

## Klasyfikacja



Kot? 1/0

## Detekcja obiektów



## Transfer stylu



<https://mc.ai/real-time-video-neural-style-transfer/>

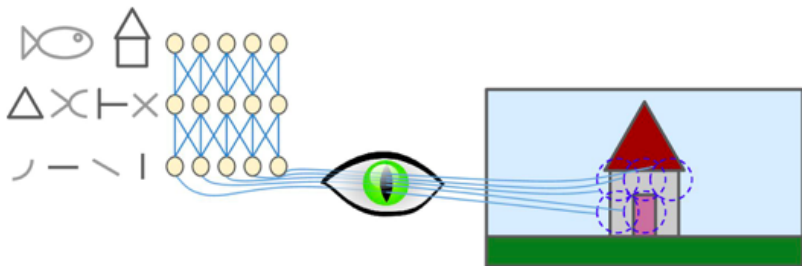
By (MTheiler) - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=75843378>

# Splotowe sieci neuronowe: wprowadzenie

- **splotowe (konwolucyjne) sieci neuronowe** (ang. *convolutional neural networks, CNN*) - wynik badań nad korą wzrokową
- używane głównie do rozpoznawania obrazów (od lat 80-tych XX wieku)
- w ciągu kilku ostatnich lat sieci CNN zdołały osiągnąć wyniki przekraczające ludzkie możliwości w pewnych skomplikowanych zadaniach, m.in. dzięki:
  - wzrostowi mocy obliczeniowych
  - ilości dostępnych danych uczących



## Inspiracja biologiczna: architektura kory wzrokowej 1/2



Aurélien Géron. Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Helion, 2018

# Inspiracja biologiczna: architektura kory wzrokowej 2/2

- Hubel & Wiesel, eksperymenty na kotach, Nagroda Nobla 1981
- wiele neuronów tworzących korę wzrokową tworzą **lokalne pola recepcyjne**, reagujące jedynie na bodźce wzrokowe mieszczące się w określonym rejonie pola wzrokowego
- pola recepcyjne poszczególnych neuronów mogą się na siebie nakładać
- pola recepcyjne łącznie tworzą całe pole wzrokowe
- pewne neurony reagują wyłącznie na obrazy składające się z linii poziomych, linii ułożonych w inny sposób itd.
- niektóre komórki nerwowe mają większe pola recepcyjne i wykrywają bardziej skomplikowane kształty
- neurony odpowiedzialne za rozpoznawanie bardziej skomplikowanych kształtów, znajdują się na wyjściu neuronów reagujących na prostsze bodźce

## Sieć z gęstymi połączeniami a sieć konwolucyjna

- podstawowa różnica: warstwy gęste uczą się cech parametrów globalnych w swoich wejściowych przestrzeniach, a warstwy konwolucyjne uczą się lokalnych wzorców
- wzorce rozpoznawane przez sieć konwolucyjną są niezależne od przesunięcia
- sieci konwolucyjne mogą uczyć się przestrzennej hierarchii wzorców

## Sieć LeNet-5

Y. LeCun, L. Bottou, Y. Bengio, P. Haffner - 1998

- znane wcześniej elementy: w pełni połączone warstwy, sigmoidalna funkcja aktywacji itd.
- nowe elementy:
  - warstwy splotowe
  - warstwy łączące

Dlaczego nie możemy stosować standardowej głębokiej sieci zawierającej w pełni połączone warstwy?

## Dlaczego nie możemy stosować standardowej głębokiej sieci zawierającej w pełni połączone warstwy?

Nie nadają się do większych obrazów z powodu olbrzymiej liczby parametrów.

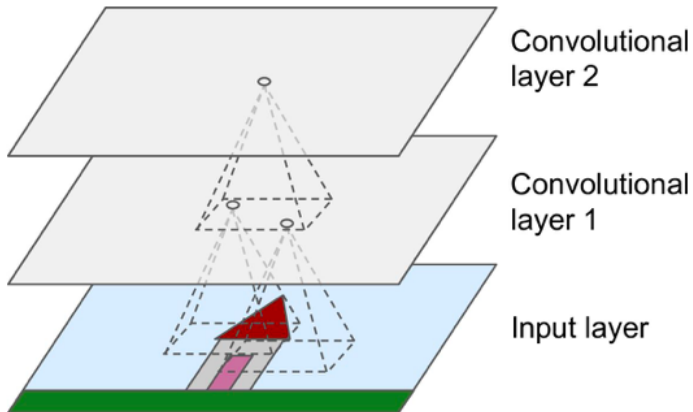
**Przykład:** obraz o rozmiarze  $100 \times 100$  zawiera 10 000 pikseli, co dla pierwszej warstwy sieci o 1000 neuronów daje łącznie 10 mln połączeń (zaledwie dla pierwszej warstwy).

W sieciach CNN problem ten zostaje rozwiązany poprzez wprowadzenie częściowo połączonych warstw.

# Warstwa splotowa

## Warstwy plotowe 1/2

Warstwa plotowa (ang. *convolutional layer*)



Aurélien Géron. *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Helion, 2018



## Warstwy splotowe 2/2

- pierwsza warstwa splotowa: neurony nie są połączone z każdym pikselem obrazu wejściowego lecz wyłącznie z pikselami znajdującymi się w ich polu recepcyjnym
- druga warstwa splotowa: każdy neuron łączy się wyłącznie z neuronami znajdującymi się w niewielkim obszarze pierwszej warstwy
- sieć może koncentrować się na ogólnych cechach w pierwszej warstwie ukrytej, łączyć je w bardziej złożone kształty w drugiej warstwie ukrytej itd.

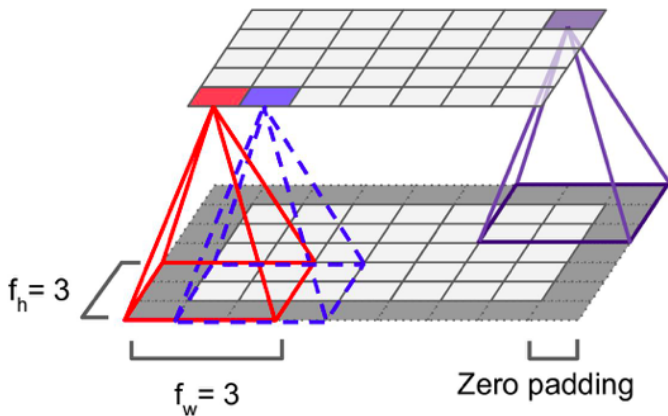
## Splot (konwolucja)

- Jako operacja matematyczna między dwoma funkcjami: mierzy całkę z ich iloczynu punktowego.
- W kontekście przetwarzania obrazów: jedna z funkcji - dwuwymiarowa macierz zawierająca wartości pikseli obrazu (zazwyczaj o dużych wymiarach), a druga z funkcji - **filtr** (zdecydowanie mniejsza macierz), nazywana **jądrem konwolucji**.
- W wyniku konwolucji obrazu z filtrem otrzymujemy nowy obraz.

# Jak uzyskać takie same wymiary każdej warstwy?

- uzupełnianie zerami
- redukcja wymiarowości

## Uzupełnianie zerami 1/2



Aurélien Géron. Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Helion, 2018

## Uzupełnianie zerami 2/2

W celu uzyskania takich samych wymiarów każdej warstwy często są dodawane zera wokół wejść

- $f_h$  - wysokość pola recepcyjnego
- $f_w$  - szerokość pola recepcyjnego
- neuron w wierszu  $i$  i kolumnie  $j$  danej warstwy jest połączony z wyjściami neuronów poprzedniej warstwy w rzędach od  $i$  do  $i + f_h - 1$  i kolumnach od  $j$  do  $j + f_w - 1$

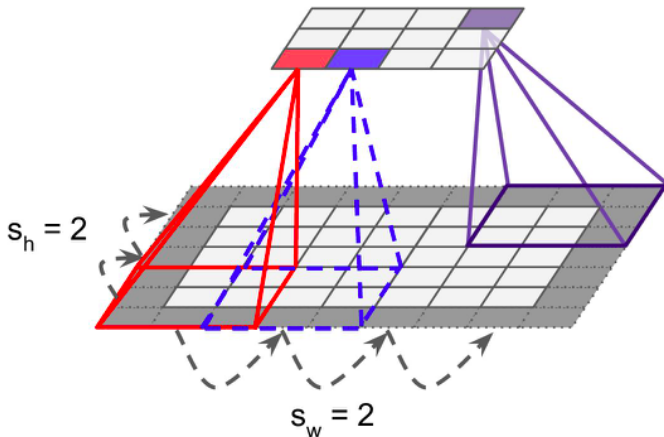
# Redukcja wymiarowości

Inny sposób: łączenie bardzo dużej warstwy wejściowej ze znacznie mniejszą kolejną warstwą poprzez rozdzielanie pól recepcyjnych

## Krok 1/2

**Krok:** odległość pomiędzy dwoma kolejnymi polami recepcyjnymi

- $s_h$  - wartość kroku w kolumnie
- $s_w$  - wartość kroku w rzędzie



## Krok 2/2

Neuron w rzędzie  $i$  i kolumnie  $j$  górnej warstwy łączy się z wyjściami neuronów dolnej warstwy w rzędach od  $i \times s_h$  do  $i \times s_h + f_h - 1$  i w kolumnach od  $j \times s_w$  do  $j \times s_w + f_w - 1$



# Filtry (jądra spłotowe)

- wagi neuronu mogą być przedstawiane jako niewielki obraz o rozmiarze pola recepcyjnego
- **filtry**: zbiory wag
- Przykład:
  - filtr pionowy:
    - macierz o wymiarach  $7 \times 7$  wypełniona zerami oprócz środkowej kolumny, która zawiera jedynki
    - neurony zawierające te wagi będą ignorować wszystkie elementy w polu recepcyjnym oprócz znajdujących się na środku pionowej linii (dane wejściowe znajdujące się poza linią przemnażane przez 0)
  - filtr poziomy: środkowa linia ułożona poziomo

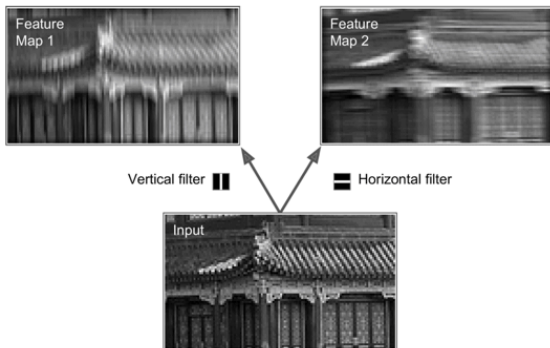
Vertical filter 

 Horizontal filter

## Mapa cech 1/2

Warstwa wypełniona neuronami wykorzystującymi ten sam filtr daje nam **mapę cech**.

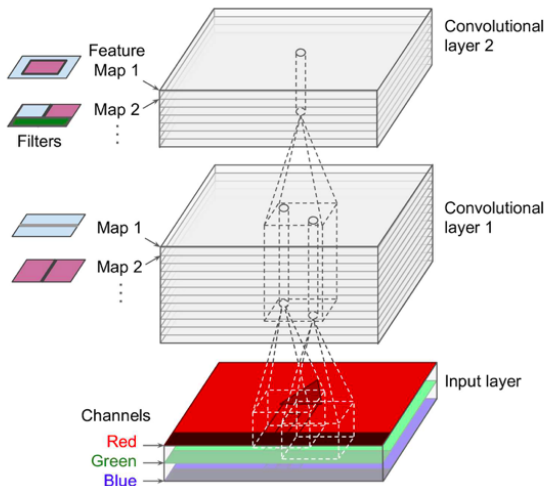
**Mapa cech** to trójwymiarowy tensor (wysokość, szerokość, głębokość/kanal).



## Mapa cech 2/2

Sieć CNN w czasie uczenia wyszukuje filtry najbardziej przydatne do danego zadania i uczy się łączyć je w bardziej złożone wzorce (np. krzyże, gdzie są aktywne oba rodzaje filtrów)

## Stosy map cech 1/2



## Stosy map cech 2/2

- Rzeczywista warstwa splotowa składa się z kilku map cech o identycznych rozmiarach
- w zakresie jednej mapy cech wszystkie neurony współdzielą te same parametry (wagi i człon obciążenia)
- inne mapy cech mogą mieć inne wartości parametrów
- pole recepcyjne neuronu nie ulega zmianie, ale "przebiega" przez wszystkie mapy cech poprzednich warstw, tj. warstwa splotowa równocześnie stosuje różne filtry na wejściach (może wykrywać wiele cech w dowolnym obszarze obrazu)

## Kanał barw / głębia

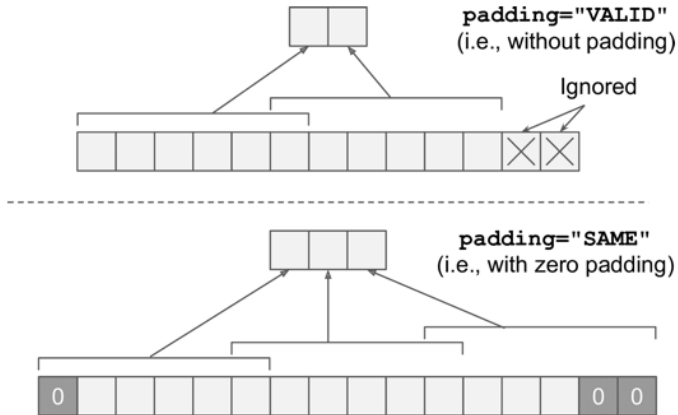
- obrazy wejściowe też składają się z kilku warstw podrzędnych po jednej na każdy **kanał barw**
- standardowo trzy kanały barw: czerwony, zielony i niebieski
- obrazy monochromatyczne: jeden wymiar głębi

## Obliczanie wartości wyjściowej neuronu w warstwie splotowej

$$z_{i,j,k} = b_k \sum_{u=0}^{f_{h-1}} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_{n'}-1} x_{i,j',k'} \cdot w_{u,v,k',k} \quad \text{gdzie} \quad \begin{cases} i' = i \times s_h + u \\ j' = j \times s_w + v \end{cases}$$

- $z_{i,j,k}$  - wyjście neuronu znajdującego się w rzędzie  $i$ , kolumnie  $j$  i mapie cech  $k$  warstwy splotowej  $l$ ,
- $s_h$  i  $s_w$  - kroki pionowy i poziomy,
- $f_h$  i  $f_w$  - wysokość i szerokość pola recepcyjnego,
- $f_n$  - liczba map cech w poprzedniej warstwie ( $l-1$ ),
- $x_{i,j,k}$  - wyjście neuronu w warstwie  $l-1$ , rzędzie  $i$ , kolumnie  $j$ , mapie cech  $k$  (lub kanale  $k$ , gdy poprzednia warstwa jest wejściową),
- $b_k$  - człon obciążenia dla mapy cech  $k$  (w warstwie  $l$ ) ("pokrętko jasności" mapy cech  $k$ ),
- $w_{u,v,k',k}$  - waga połączenia pomiędzy dowolnym neuronem w mapie cech  $k$  warstwy  $l$  a jego wejściem mieszczącym się w wierszu  $u$ , kolumnie  $v$  (względem pola recepcyjnego neuronu) a mapą cech  $k$

## Opcje uzupełniania zerami 1/2





## Opcje uzupełniania zerami 2/2

- VALID: warstwa spłotowa nie używa uzupełniania zerami oraz może ignorować niektóre rzędy i kolumny, w zależności od rozmiaru kroku
- SAME:
  - uzupełnianie zerami stosowane w razie potrzeby
  - liczba neuronów wyjściowych jest równa zaokrąglonemu w górę ilorazowi neuronów wejściowych i rozmiaru kroku (np.  $13/5 \approx 3$ )
  - następnie dodawane zera wokół danych wejściowych żeby wyszło jak ""najrówniej""

# Warstwa łącząca

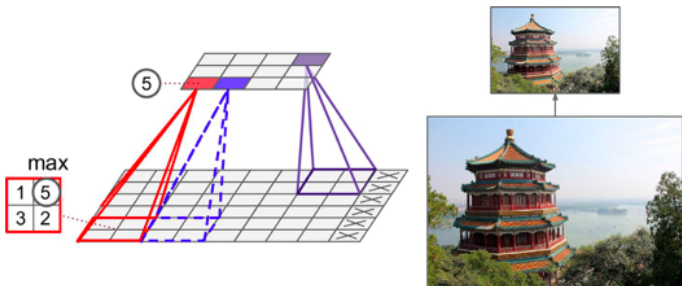
## Warstwa łącząca

Celem **warstwy łączącej** (ang. *pooling layer*) jest **podpróbkiwanie** (zmniejszenie) obrazu wejściowego w celu zredukowania obciążenia obliczeniowego, wykorzystania pamięci i liczby parametrów

- każdy neuron stanowiący część warstwy łączącej łączy się z wyjściami określonej liczby neuronów warstwy poprzedniej, mieszczącej się w obszarze niewielkiego pola recepcyjnego (podobnie jak w przypadku warstw splotowych)
- definiujemy rozmiar pola, wartość kroku, rodzaj wypełniania zerami, itd. (podobnie jak w przypadku warstw splotowych)
- warstwa łącząca **nie zawiera żadnych wag** - jej jedynym zadaniem jest gromadzenie danych za pomocą jakiejś funkcji agregacyjnej (np. maksymalizującej, uśredniającej)

## Maksymalizująca warstwa łącząca (ang. *max pooling layer*)

jądro łączące: rozmiar  $2 \times 2$ , krok: 2, brak uzupełniania zerami

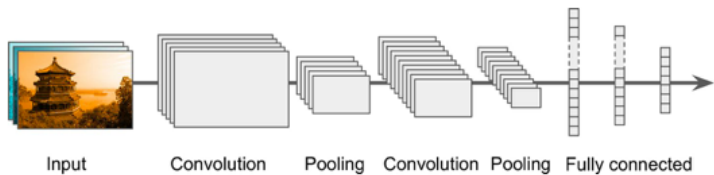


Jedynie maksymalna wartość z każdego jądra zostaje przekazana do następnej warstwy.

Aurélien Géron. *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Helion, 2018

# Architektury splotowych sieci neuronowych

# Typowa architektura sieci CNN



Aurélien Géron. *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Helion, 2018

# Wyzwanie ImageNet 1/2

## *Large Scale Visual Recognition Challenge*

**Przykładowe zadanie:** 150 tyś. obrazów ręcznie etykietowanych na obecność obiektów z predefiniowanej hierarchii 1000 kategorii.

**Cel:** podanie listy 5 kategorii, których obiekty występują na zadanym obrazie.

## Wyzwanie ImageNet 2/2

### Large Scale Visual Recognition Challenge

Wyniki:

rok	błędy
człowiek	≈ 5%
2010	28,2%
2011	25,8%
2012	16,4%
2013	11,7%
2014	6,7%
2015	3,6%
2016	3,0%
2017	2,3%



# Architektury spłotowych sieci neuronowych

- LeNet-5 (1998)
- AlexNet (2012)
- GoogLeNet (2014)
- ResNet (2015)

# Literatura

- ① Aurélien Géron. *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Helion, 2018
- ② François Chollet. *Deep learning. Praca z językiem Python i biblioteką Keras*. Helion, 2019

Dziękuję za uwagę!

# Pytania testowe

Ponieważ warstwy łączące nie mają parametrów (wag), nie wpływają one na obliczenia w ramach propagacji wstecznej:

- prawda
- fałsz

Zaletą korzystania ze spłotowych sieci neuronowych (CNN) jest "współdzielenie parametrów". Które z poniższych stwierdzeń dotyczących współdzielenia parametrów w sieciach CNN jest prawdziwe?

- ❑ Zmniejsza całkowitą liczbę parametrów, zmniejszając w ten sposób nadmierne dopasowanie modelu
- ❑ Pozwala na użycie detektora cechy w wielu lokalizacjach w całym obrazie wejściowym
- ❑ Umożliwia algorytmowi spadku gradientu ustawienie wielu parametrów na zero, dzięki czemu połączenia są rzadkie
- ❑ Umożliwia współdzielenie parametrów wyuczonych dla jednego zadania w innym zadaniu (*transfer learning*)

## Zaletą korzystania z warstw splotowych jest "rzadkość połączeń". Co to znaczy?

- Każda aktywacja w kolejnej warstwie zależy jedynie od małej liczby aktywacji z poprzedniej warstwy
- Każda warstwa w sieci konwolucyjnej jest połączona tylko z dwoma innymi warstwami
- Regularyzacja powoduje, że algorytm spadku gradientu ustawia wartości wielu parametrów na zero
- Każdy filtr jest połączony z każdym kanałem z poprzedniej warstwy